Measuring And Calculating Balancing Authority Frequency Response

In Interconnections With More Than One Balancing Authority

Mike Potishnak

April 27, 2011

Introduction

The NERC Frequency Responsive Reserve Standard Drafting Team (FRRSDT) has been evaluating methods for measuring and calculating Balancing Authority (BA) frequency response for Interconnections with more than one BA. Empirical studies conducted thus far have addressed sampling interval and averaging technique selection, and data quality concerns that may influence the ultimate selection of these techniques.

The FRRSDT's preliminary choice for the sampling interval uses the point A and B computational specifications as shown in Table 1, and is referred to herein as the "20 to 52 second" metric. This sampling interval and other sampling intervals

Table 1

20 to 52 Second Sampling Metric Specification

Definitions of Frequency Values for Frequency Response Calculation			
Scan Rate	**T 0 Scan**	**A Value (average)**	**B Value (average)**
6-Seconds	**Identify first significant change in frequency as the T 0 scan**	**Average of T-1 through T-2 scans**	**Average of T+4 through T+8 scans**
5-Seconds		**Average of T-1 through T-2 scans**	**Average of T+5 through T+10 scans**
4-Seconds		**Average of T-1 through T-3 scans**	**Average of T+6 through T+12 scans**
3-Seconds		**Average of T-1 through T-5 scans**	**Average of T+7 through T+17 scans**
2-Seconds		**Average of T-1 through T-8 scans**	**Average of T+10 through T+26 scans**

were evaluated in these studies. The FRRSDT developed PI and EXCEL interfaces for each of the specified scan rates, and distributed them to BAs volunteering to provide data to support this effort. Seventy frequency events in 2009 were chosen for sampling, based mostly on selections made from the ISO New England high speed frequency data recorder, though some other samples were received from another BA to cover 2 intervals with ISO New England data unavailability. These events are shown in Appendix 1. Six Eastern Interconnection BAs provided data to support this effort, but the data from one BA could not be used due to data quality issues that have yet to be resolved. All calculations used actual net interchange data from the EMS data stored in PI. The frequency values of the selected events are shown in Appendix 1. The 12 to 20 second metric and the best ANI value in 0 to 20 seconds" metric used the frequency values shown in Appendix 1 for points A and B. Other metrics used the frequency values archived that originated in the EMS.

Overview Of The Empirical Studies Performed

The studies described herein were conducted during 2010 and in the first four months of 2011. The results were discussed during meetings and conference calls with the FRRSDT during this period and led to additional analyses. In particular, comments received from the industry after the initial posting of the proposed standard (BAL-003-1) sought clarification on the FRRSDT's choice of a median (instead of an average or linear regression) as the method of combining individual samples to obtain the ultimate frequency response performance result.

To promote data confidentiality, results have been shared as normalized values without a direct association with the specific BA (e.g., BA #5 instead of ISO New England).

1. Sampling Intervals

Four sampling intervals were evaluated in the studies. The 20 to 52 second metric described in Table 1 was evaluated. A 12 to 20 second metric, which is typical of what Balancing Authorities have used for frequency bias sampling in the past, was also evaluated. To explore the differences in observations in these two metrics, a 20 to 40 second metric was also evaluated. A "best ANI value in 0 to 20 seconds" was also included, mainly to determine the extent to which BAs could “cherry pick” ANI values to inflate the computed frequency response.

2. Selection of Time Zero

In the course of evaluating the results from different sampling intervals, it became apparent that the selection of time zero influences the ultimate scores. As frequency is usually varying a small amount in some direction prior to the contingency that caused the frequency event, it is not always clear when the event actually begins. Given that each BA would need to use discretion in this choice without further guidance, the use of the largest cross-scan drop in frequency as time zero was compared with the first cross scan drop above some threshold. The largest drop removes discretion, but it might be worth the additional complexity to select a minimum downward deviation as time zero.

3. Point A Averaging Interval

In the course of evaluating the results from different sampling intervals, some data skew was observed in which the archived data would reflect a large change in actual net interchange prior to a change in frequency. To evaluate this data quality issue, the sampling specified for point A in Table 1 above was modified experimentally to skip one and then 2 samples immediately preceding time zero.

4. Sorting Criteria

To explore the sensitivity of the computed frequency response for various sample sets, the 70 samples were subdivided based on the following criteria: elimination of samples in which the BA was contingent; on and off peak with contingencies removed; point B less than 59.95 Hz with contingencies removed; and, (2) and (3) above.

5. Averaging Techniques

Mean, median, and linear regression were evaluated. As linear regression was added to the analysis later in these studies, it was only used with the 20 to 52 second metric.

6. Convergence

Subject matter experts have shared their somewhat anecdotal observations that the frequency response calculations that they performed would converge to a reasonably stable value with about 20 samples. The data available to the FRRSDT was used to evaluate the convergence properties of the frequency bias calculation process.

7. Resiliency To Contamination Of Actual Net Interchange Values

The actual net interchange values used in BA frequency response sampling is influenced by other operating phenomena. The next section describes those influences in detail. In this study, contaminated values are removed by subject matter experts, and the results from the use of a mean, median, and linear regression are compared to see which metric is more resilient given the high probability of many BAs have contaminants in their actual net interchange data. This evaluation was performed very late in this overall effort, to further support the FRRSDT's decision-making process in picking the method to use to combine the individual frequency response samples for an overall score, and also to provide a quantitative response to industry comments.

8. Supplemental Regulation

A limited study of the impact of supplemental regulation on one BA's frequency response was performed, to determine the sensitivity of the results to its inclusion in and exclusion from actual net interchange values.

Data Quality Concerns Related To The Use Of The Actual Net Interchange Value

Actual net interchange for a typical BA is the summation of its tie lines to other BAs. In some cases, there are pseudo-ties in it which reflect the effective removal of load and/or generation from another BA, or it could include supplemental regulation as well. But in the typical scenario, actual net interchange values that are extracted from EMS data archiving can be influenced by data latency times in the data acquisition process, and also any timestamp skewing in the archival process. The point A Averaging Interval discussion above reflects some of this reality.

Of greater concern, however, are the inevitable variations of other operating phenomena concurrent within a frequency event. The impacts of these phenomena are superimposed on actual net interchange values along with the frequency response that we wish to measure through the use of the actual net interchange value.

To explore this issue further, let's begin with the idealized condition:

- frequency is fairly stable at some value near or a little below 60 Hz
- ACE of the non-contingent BA of interest is 0 and has been 0 for an extended period, and AGC control signals have not been issued recently
- Actual net interchange is "on schedule", and there are no schedule changes in the immediate future
- BA load is flat
- All generators not providing AGC are at their targets
- Variable generation such as wind and solar are not varying
- Operators have not directed any manual movements of generation recently

And when the contingency occurs in this idealized state, the change in actual net interchange will be measuring only the decline in load due to lesser frequency and generator governor response, and, none of the contaminating influences. While the ACE may become negative due to the actual frequency response being less than that called for by the frequency bias setting within the BA's AGC system, this contaminating influence on measuring frequency response will not appear in the actual net interchange value if the measurement interval ends before the generation on AGC responds.

Now let's explore the sensitivity of the resultant frequency bias sample to the relaxation of these idealized circumstances.

1. The "60 Hz load" increases moderately due to time of day concurrent with the frequency event. If the frequency event happens before AGC or operator-directed manual load adjustments occur, then the actual net interchange will be reduced and the frequency response will be underestimated. But if the frequency event happens while AGC response and/or manual adjustments occur, then the actual net interchange will be increased and the frequency response will be overestimated.
2. The "60 Hz load" decreases moderately due to time of day concurrent with the frequency event. If the frequency event happens before AGC or operator-directed manual load adjustments occur, then the actual net interchange will be increased and the frequency response will be overestimated. But if the frequency event happens while AGC response and/or manual adjustments occur, then the actual net interchange will be decreased and the frequency response will be underestimated.
3. In anticipation of increasing load during the next hour, the operator increases manual generation before the load actually appears. If the frequency event happens while the generation "leading" the load is increasing, then the actual net interchange will be increased and the frequency response will be overestimated. But if the frequency event occurs when the result of AGC signals sent to offset the operator's leading actions, then the actual net interchange will be decreased and the frequency response is underestimated.
4. In anticipation of decreasing load during the next hour, the operator decreases manual generation before the load actually declines. If the frequency event happens while the generation "leading" the load downward is decreasing, then the actual net interchange will be decreased

and the frequency response will be underestimated. But if the frequency event occurs when the result of AGC signals sent to offset the operator's leading actions, then the actual net interchange will be increased and the frequency response is overestimated.

5. A schedule change to export more energy is made at 5 minutes before the top of the hour. The BA's "60 Hz load" is not changing. The schedule change is small enough that the operator is relying on upward movement of generators on AGC to provide the additional energy to be exported. The time at which the AGC generators actually begin to provide the additional energy is dependent on how much time passes before the AGC algorithm gets out of its deadbands, the individual generator control errors gets large enough for sending out the control signal, and maybe 20 seconds to 3 minutes for the response to be effected. The key point here is that it is not clear when the effects of a schedule change, as manifested in a change in generation and then ultimately a change in actual net interchange, will occur. Since it seems implausible that most BAs would have the expertise, readily available data, and possibly time and interest to produce an "accurate" correction to actual net interchange for affected frequency bias samples, ***it is highly recommended that BA's not be allowed to provide corrections for changes in scheduled net interchange****.* This exercise reinforces the desirability of not selecting samples during periods of probable substantive schedule changes, as has been stated previously in the event selection criteria produced by the FRRSDT.
6. With the expected penetration of wind in the near future, unanticipated changes in their output will tend to affect actual net interchange and add noise to the frequency response observation process.

To a greater or lesser extent, 1 through 4 above are happening continuously for the most part with most BAs in the Eastern and Western Interconnections. The frequency response is buried within the typical hour to hour operational cacophony superimposed on actual net interchange values. The choice of metrics will be important to artfully extract frequency response from the noise.

In dealing with the noise in frequency response measurements, there may be a tendency to assume that with a sufficient sample size the noise will balance out. Addressing that assumption, even if there was a sample set size in excess of 1000, is it necessarily true that those values contaminated by the largest values emanating from 1 through 4 above would be equal by head count and also equal by magnitude? In the discussion to follow, the 70 sample data set will be evaluated for symmetry in head count of values that are obviously contaminated by the phenomena described in 1 through 4 above. If a sample set size of 25 is used, does the empirical data indicate that the noise is balancing out, or can be expected to more or less always balance out?

Review of the Empirical Results

In Table 2, the results for the "best ANI value in 0 to 20 seconds", 12 to 20 second metric, and the 20 to 52 second metrics are provided for BAs 1, 2, 3, and 4. The results in Table 2 make use of the largest frequency change observed as "time zero." Median values, average values, and the linear regression value (for the 20-52 second metric only) are provided for 5 scenarios, and the sample counts in each scenario are shown on the far right. Scenario 1 includes all samples available. Note that the initial sample set size was 71, but BA's #3 and #4 had one sample for which no data was available. (In later studies, a consistent set of 70 samples is used by dropping the final sample for BA#1 and BA#2.) Scenario #2 eliminates those samples whose change in actual net interchange was uncharacteristically large. For BA#1, 3 of the 4 samples eliminated were due to the fact that it was the contingent BA, while BA#2 had 2 of its contingencies eliminated. BA#3 and BA#4 did not suffer any contingencies among the sample set, and did not have an uncharacteristically large change in actual net interchange in any sample. (Note that this is a very coarse screening that is refined in a study cited later.)

Scenarios 3, 4, and 5 provides further sorting of scenario 2. Scenario 3 only includes those samples found in scenario 2 for which point B is 59.95 Hz or lower. Scenario 4 has only on peak events found in scenario 2. Scenario 5 has only off peak events found in scenario 2.

The 12-20 second median value of scenario 3 is used to normalize the results to preserve data confidentiality. Normalized values are computed by dividing each MW/.1 Hz value by the 12-20 second median value of scenario 3 and then multiplying by 100%.

Table 2

Comparison of Metrics When T=0 is the Largest Frequency Drop

		Normalized median values			Normalized average values				
2009 BA#1 frequency response		median 12-20 sec <59.95 Hz as base			median 12-20 sec <59.95 Hz as base				
scenario #		0-20 sec best MW/.1 Hz	12-20 sec average MW/ .1 Hz	20-52 sec MW/ .1 Hz	0-20 sec best MW/.1 Hz	12-20 sec average MW/ .1 Hz	20-52 sec MW/ .1 Hz	linear regression	# of samples
1	all samples	182	96	105	125	30	40	71	71
2	drop large tie changes	189	100	109	195	104	123	109	67
3	drop large tie changes and point B<59.95	193	100	96	214	117	141	117	27
4	drop large tie changes, on peak only	189	101	105	200	106	114	104	46
5	drop large tie changes, off peak only	173	96	111	168	96	140	120	21
2009 BA#2 frequency response									
scenario #									
1	all samples	133	86	135	123	28	135	109	71
2	drop large tie changes	134	95	135	125	45	144	137	69
3	drop large tie changes and point B<59.95	131	100	135	119	41	143	136	28
4	drop large tie changes, on peak only	129	76	118	121	19	134	128	46
5	drop large tie changes, off peak only	134	90	136	127	92	160	152	23
2009 BA#3 frequency response									
scenario #									
1	all samples	174	104	72	175	109	81	78	70
2	drop large tie changes	174	104	72	175	109	81	78	70
3	drop large tie changes and point B<59.95	171	100	68	177	110	78	73	31
4	drop large tie changes, on peak only	176	104	70	177	106	82	81	49
5	drop large tie changes, off peak only	150	100	83	166	113	81	73	21
2009 BA#4 frequency response									
scenario #									
1	all samples	127	86	100	139	83	98	99	70
2	drop large tie changes	127	86	100	139	83	98	99	70
3	drop large tie changes and point B<59.95	140	100	107	161	96	106	104	28
4	drop large tie changes, on peak only	132	86	101	146	87	101	100	48
5	drop large tie changes, off peak only	120	80	95	123	77	90	96	22

In reviewing the results presented in Table 2, the following observations are of interest:

- For BAs 1, 3, and 4, the best 0-20 second value, either its mean or median, is much higher than any other metric in all scenarios. If BAs are not given a consistent methodology to compute the point B actual net interchange value, then some BAs could "cherry pick" within the samples. They could also "cherry pick" among samples to further inflate their score. ***It is conclusive that all BAs need to use a common methodology to avoid "cherry picking".***
- When BA#1 and BA#2's results are compared across the first two scenarios, the median value of scenario 1 is consistently closer to its value in scenario 2 than either the mean or linear regression value. (Since BAs 3 and 4 did not have any contingencies in the sample set, nor extremely large changes in actual net interchange, scenarios 1 and 2 are inherently the same.) This is an indicator that the median will be more resilient to data quality problems, which are not just limited to a BA being the contingent BA. Resiliency to data quality problems will be explored in depth below.
- For BAs 1, 2, and 4, the 20 to 52 second results are consistently higher than the 12 to 20 second metric. Upon discussion with subject matter experts for BAs 1 and 2, the difference reflects influences from some hydro generation providing actual AGC response starting in about 25 seconds after the event, which has caused the ACE to deflect substantially negative due to the difference between the frequency bias setting in AGC and the actual frequency response. Also included therein is some legitimate governor response from generators with high, intermediate, and low pressure turbines. The steam transit time to get an additional "boost" in output from the low pressure turbine can approach 50 seconds. ***It is conclusive that some AGC action will appear as frequency response with the use of the 20 to 52 second metric.***
- For BA#3, the 20 to 52 second metric values are substantially lower than their 12 to 20 second counterparts. Their subject matter experts strongly suspect that the difference reflects squelched response at some of their

generating stations, but the suspicion has yet to be confirmed. ***While more analysis is needed for many more BAs, this may be an indicator that the 20 to 52 second metric will flag squelched response.***

- A comparison of on and off peak results does not yield a common pattern among the four BAs. BA1 values drop 5 to 10% from on to off peak for the 12 to 20 second metric, while the 20 to 52 second metric values increased by 6 to 26%. This difference may be attributable to the difference in generation mix that is not at full load. BA#2's frequency response increases somewhat during off peak periods, while BA#4 goes down, and BA#3's changes are small and differ directionally among its metrics. ***While more data is needed for more BAs, there does not seem to be a need to create special metrics based on the peak period, nor should sampling be restricted to on peak periods as was done in prior practices.***
- In comparing linear regression results with the median, the largest differences are in scenario 1. However, the differences between median and linear regression in the other scenarios is usually not very large. This will be addressed in greater detail later.
- When comparing mean and median in scenario 1, there are major differences for BAs 1 and 2, but minor differences for BAs 3 and 4. BA2 mean and median differ very significantly for the 12 to 20 second metric in scenarios 2, 3, and 4, but less so in scenario 5. It is suspected that the on peak samples for BA 2 have some heavily contaminated values and the median circumvents their influence. This will be addressed in greater detail later.

Limited comparisons were made to determine the sensitivity of the different frequency response metrics when results similar to Table 2 above are developed using the first frequency drop above 5 mHz as time zero. A comparison of BA 1's results for scenario 2 is shown in Table 3. Basically, the scores show improvement in the range of 4 to 8%. For BA1, only 7 out of 67 samples had a different time zero when the first frequency drop above 5 mHz was used as the criterion for

Table 3

Using Largest Frequency Change Vs. First Change >= 5 mHz for BA #1

BA # 1		normalized	normalized
scenario 2		frequency	frequency
		response	response
		time zero=	time zero=
averaging	sampling	largest	first change
technique	interval	change	>= 5 mHz
median	12-20 seconds	100	108
median	20-52 seconds	109	116
mean	12-20 seconds	104	111
mean	20-52 seconds	123	128
regression	20-52 seconds	109	113

Table 4

Using Largest Frequency Change Vs. First Change >= 5 mHz for BA #2 & 3

			median	median	median	median	average	average	average	average	regression	regression
			largest	first >5 mHz	largest	first >5 mHz	largest	first >5 mHz	largest	first >5 mHz	largest	first >5 mHz
			frequency	frequency	frequency	frequency	frequency	frequency	frequency	frequency	frequency	frequency
			drop	drop	drop	drop	drop	drop	drop	drop	drop	drop
			12-20 sec	12-20 sec			12-20 sec	12-20 sec				
			average	average	20-52 sec	20-52 sec	average	average	20-52 sec	20-52 sec	20-52 sec	20-52 sec
BA #		scenario	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/ .1 Hz
2	1	all samples	86	58	135	119	28	16	135	123	109	99
2	2	drop large tie changes	95	61	135	119	45	27	144	131	137	124
2	3	drop large tie changes and point B<59.95	100	67	135	128	41	37	143	126	136	118
2	4	drop large tie changes, on peak only	76	63	118	100	19	13	134	119	128	110
2	5	drop large tie changes, off peak only	90	57	136	131	92	57	160	151	152	146
3	1	all samples	104	131	72	81	109	128	81	84	78	81
3	2	drop large tie changes	104	131	72	81	109	128	81	84	78	81
3	3	drop large tie changes and point B<59.95	100	137	68	76	110	129	78	81	73	75
3	4	drop large tie changes, on peak only	104	132	70	81	106	127	82	84	81	83
3	5	drop large tie changes, off peak only	100	114	83	83	113	130	81	82	73	75

time zero instead of the largest frequency drop. Results for BA 2 and BA 3 are shown in Table 4 for all scenarios. Curiously, BA 2 scores decline noticeably using the first frequency drop, while BA 3 scores improve noticeably. BA 2 scores probably drop because some of the sampling periods begin 1 scan earlier and thus 1 scan of AGC response is excluded. BA 3 scores probably improve because the earlier start for some of the sampling periods makes the squelched response less observable. ***Summarizing, the results are sensitive to the method used to choose time zero – largest is easier to administer, but more research is needed in this area.***

A comparison of scenarios 2 and 3 in Table 2 illustrate the sensitivity to point B values being above or below 59.95 Hz. BA# 5 was not included in table 2 above because its data was only collected using the first frequency drop greater than 5 mHz. BA 5 results in their entirety are shown in Table 5. The change in results for all 5 BAs are shown in Table 6, with BA 1 through 4's data originating in Table 2, and BA 5's data originating in Table 5.

Table 5

Comparison of Metrics When T=0 is the First Frequency Drop >= 5 mHz for BA #5

		Normalized median values			Normalized average values				
2009 BA#5 frequency response		median 12-20 sec <59.95 Hz as base			median 12-20 sec <59.95 Hz as base				
		0-20 sec	12-20 sec		0-20 sec	12-20 sec			
scenario		best	average	20-52 sec	best	average	20-52 sec	linear	# of
#		MW/.1 Hz	MW/ .1 Hz	MW/ .1 Hz	MW/.1 Hz	MW/ .1 Hz	MW/ .1 Hz	regression	samples
1	all samples	61	48	118	96	23	88	122	71
2	drop large tie changes	73	52	114	97	57	87	97	70
3	drop large tie changes and point B<59.95	121	100	120	133	81	91	109	25
4	drop large tie changes, on peak only	58	42	100	106	59	93	109	48
5	drop large tie changes, off peak only	86	65	120	79	51	74	74	22

Table 6

Percent Change In Calculated Frequency Response When Events With Point B > 59.95 Are Removed (Scenario 2 vs. Scenario 3)

averaging technique	sampling interval	BA #1	BA #2	BA #3	BA #4	BA #5
median	12-20 seconds	0	5	-4	14	48
median	20-52 seconds	-13	0	-4	7	6
mean	12-20 seconds	13	-4	1	13	24
mean	20-52 seconds	18	-1	-3	8	4
regression	20-52 seconds	8	-1	-5	5	12

In reviewing the results shown in Table 6, BAs 2 and 3 seem insensitive to the application of the 59.95 Hz constraint, while BA 1's score is erratic in response to it. BA 4 has a moderate increase among the metrics, while BA #5 shows a large increase in 2 of the metrics. If changes below 10% are assumed to be noise, then 7 out of 8 remaining measures in Table 6 show an increase when the events with point B above 59.95 Hz are removed. ***In summary, it seems that events with point B above 59.95 Hz will result in lower scores, but a larger set of BAs are needed to make any firmer quantitative statements. It may make sense to have the majority of frequency events having a point B below 59.95 Hz.***

One of the BAs providing data saw about a 20% improvement in its 2010 results by choosing its sampling of point A to start and end one scan earlier than specified, while using a 20 to 40 second sampling interval with a median and a mean. This BA's score improved by about 5% for its 2009 median value and 15% in its average value for the 20 to 40 second metric by shifting it sampling of point A one scan earlier. Interestingly, this BA's score changed minimally for the 12 to 20 second and 20 to 52 second metric with the mean and median in 2009, and the linear regression value in 2009 was unchanged as well. However, 3 other BAs experienced very little or no change in results with a 20 to 40 second sampling interval with a median and a mean, having point A start and end one scan earlier, using 2009 samples. ***In summary, more analysis is needed to evaluate the impact of time skew related to the point A sampling interval.***

All BAs except BA5 were evaluated using a 20 to 40 second metric later in the process. The elimination of the latest 12 seconds of sampling was an attempt to remove some of the AGC influence from the metric. When the analysis was performed for BA2 and BA3, it was normalized based on the first frequency drop >= 5 mHz data for the 12 to 20 second mean, and therefore could not be included in Table 2 above. In Table 7, BA1 and BA4 are being normalized based on the largest frequency drop. This inconsistency is of little importance as long as the data for BA2 and BA3 in Table 7 are not compared with Table 2.

For BA1 and BA2, the 20 to 40 second sampling interval in scenarios 2 through 5 often produce a measure with a somewhat larger result than the 20 to 52 second sampling interval. This is an indicator that lopping off 12 seconds at the back end of point B will not eliminate AGC influences. The decline that appears when extending from 40 to 52 seconds may be evidence of some squelched frequency response. BA3's scores are essentially the same at the 20 to 40 and 20 to 52 second sampling intervals, which may indicate that the suspected squelched response will still be observed between 20 and 40 seconds. BA4's response for 20 to 40 seconds is in between the 12 to 20 and 20 to 52 second response, as might be expected. This may reflect this BAs hydro frequency response showing up between 40 and 52 seconds. The 20 to 40 second sampling interval is not clearly superior to the other sampling intervals. The author speculates, with gut instinct but without any empirical validation, that a 15 to 30 second sampling interval would be the best. ***In summary, many BAs need to provide scan rate data from 60 seconds before to 90 seconds after the frequency event to support an analysis to determine the best (or maybe just the least evil) sampling interval.***

Table 7

Comparison of a 20-40 Second Metric With The 12-20 Second Metric

And The 20-52 Second Metric

BA#	scenario #		median of 12-20 sec average MW/ .1 Hz	median of 20-40 sec average MW/ .1 Hz	median of 20-52 sec average MW/ .1 Hz		average of 12-20 sec average MW/ .1 Hz	average of 20-40 sec average MW/ .1 Hz	average of 20-52 sec average MW/ .1 Hz	regressior 20-52 sec average MW/ .1 Hz
1	1	all samples	96	104	105		30	40	36	71
1	2	drop large tie changes	100	106	109		104	123	117	109
1	3	drop large tie changes and point B<59.95	100	114	96		117	141	137	117
1	4	drop large tie changes, on peak only	101	105	105		106	114	110	104
1	5	drop large tie changes, off peak only	96	126	111		96	140	130	120
2	1	all samples	110	195	199		64	172	193	150
2	2	drop large tie changes	134	210	204		136	218	218	209
2	3	drop large tie changes and point B<59.95	100	214	202		85	176	200	189
2	4	drop large tie changes, on peak only	123	187	184		130	207	208	200
2	5	drop large tie changes, off peak only	177	243	212		159	235	226	223
3	1	all samples	96	58	56		93	61	60	58
3	2	drop large tie changes	96	58	56		93	61	60	58
3	3	drop large tie changes and point B<59.95	100	53	54		95	60	58	54
3	4	drop large tie changes, on peak only	96	58	56		92	63	61	60
3	5	drop large tie changes, off peak only	83	61	58		91	57	55	53
4	1	all samples	86	93	100		83	90	98	99
4	2	drop large tie changes	86	93	100		83	90	98	99
4	3	drop large tie changes and point B<59.95	100	103	107		96	98	106	104
4	4	drop large tie changes, on peak only	86	97	101		87	97	101	100
4	5	drop large tie changes, off peak only	80	90	95		77	84	90	96

As stated above, the other operating phenomena at times will result in an unacceptable contamination of the actual net interchange values. Based on discussions with subject matter experts in these BAs, a negative value (when non-contingent, and implying that energy will enter and not exit the BA area) is clearly a wrong value. In Table 8, the number of negative values for scenarios 2 (drop large errors only) and 3 (drop large errors and samples with point B> 59.95 Hz) are shown for all 5 BAs. The data shown is for time zero being the largest frequency drop for all BAs except for BA 5. Aside from the contaminated values flagged by being less than 0, it is apparent that there are plenty of values slightly above 0 that are obviously contaminated as well when the actual sample values are reviewed. ***In summary, it is impractical to rely on a single sample to be above some threshold value, and frequency response values will be meaningful ONLY when used with some type of averaging technique, and the choice of averaging technique will impact the accuracy of the score, given the data quality problems associated with using actual net interchange values.***

Table 8

Head Count of Samples Indicating Energy Sinking In Non-Contingent BAs

scenario	2= all except large errors			
scenario	3 =no large errors & <59.95 Hz			
		12-20 seconds	20-52 seconds	
		# negative	# negative	
BA#	scenario	values	values	# samples
1	2	6	5	67
1	3	1	2	27
2	2	16	4	67
2	3	5	2	28
3	2	0	7	70
3	3	0	4	31
4	2	3	1	70
4	3	1	0	28
5	2	19	18	70
5	3	5	6	25

The FRRSDT began its work using the opinion of subject matter experts that 25 samples will converge to a result that is very similar to that found if the sample count is doubled, for example. In Figures 1.1 through 1.5 below, the frequency bias samples for BA1 through BA5 were used to evaluate this hypothesis. The "sliding sample count averages" are shown for the mean and median values of the 12 to 20 and 20 to 52 second sampling intervals. The number of samples is shown on the x-axis, and the normalized values of the metrics are shown on the y-axis. The 70 sample set was reduced to 33 samples by skipping every other sample in chronological order. For example, the y-values shown for the x-axis value of 3 includes 3 samples in the result, while y-values shown for the x-axis value of 33 includes 33 samples in the result. In all cases in Figures 1 through 5, a reasonably stable value is achieved (though at differing values) by 25 samples, and reasonable convergence is shown at 20 samples as well. ***In summary, 25 samples should be a sufficient sample set size, provided that the right choices are made for the sampling interval and the averaging technique.***

Figure 1.1

BA-1 Convergence Properties for the Mean and Median for the 12-20 Second And 20-52 Second Sampling Intervals

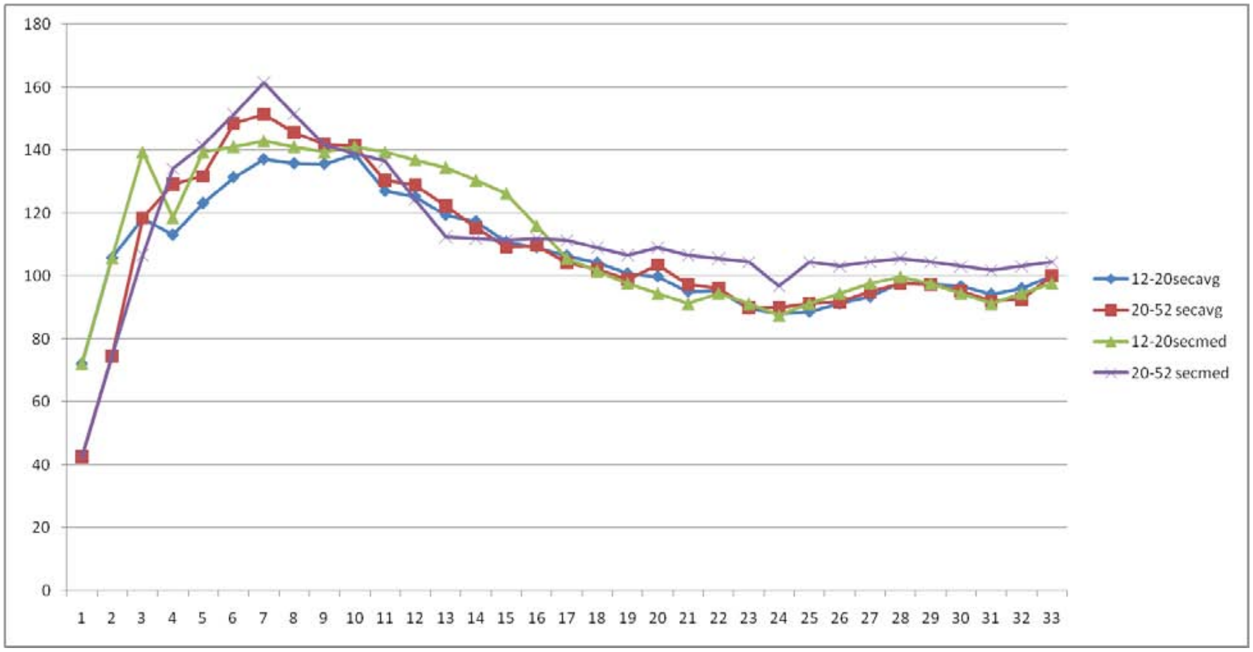


Figure 1.2

BA-2 Convergence Properties for the Mean and Median for the 12-20 Second And 20-52 Second Sampling Intervals

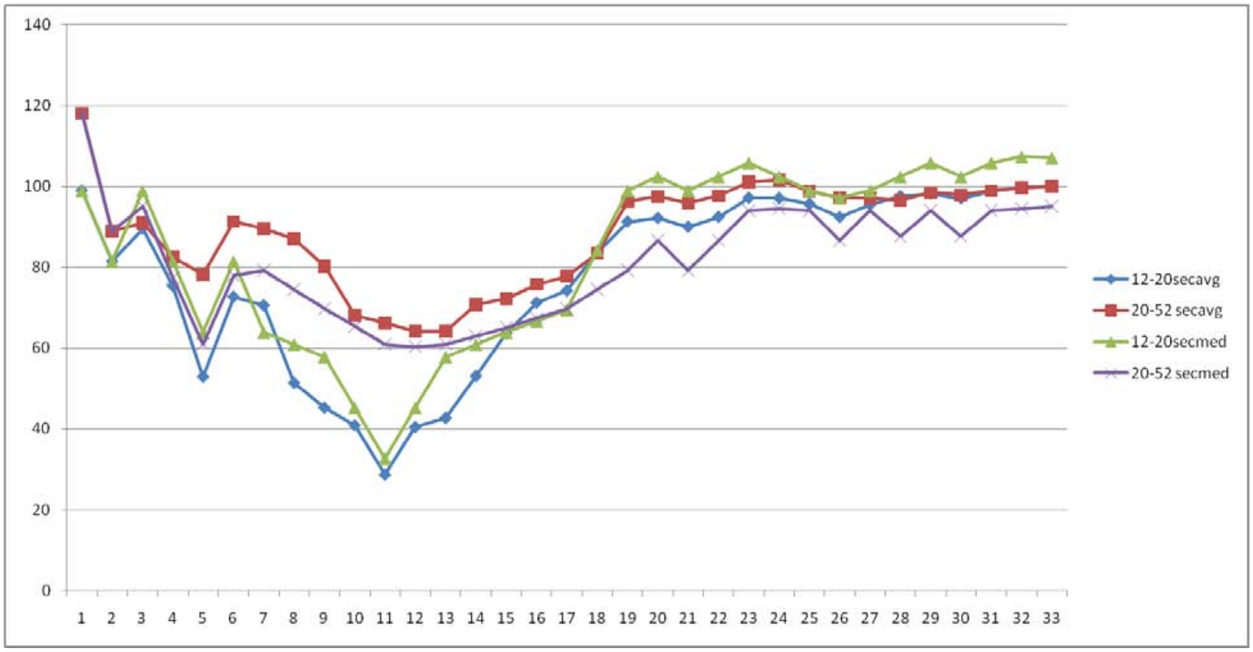


Figure 1.3

BA-3 Convergence Properties for the Mean and Median for the 12-20 Second And 20-52 Second Sampling Intervals

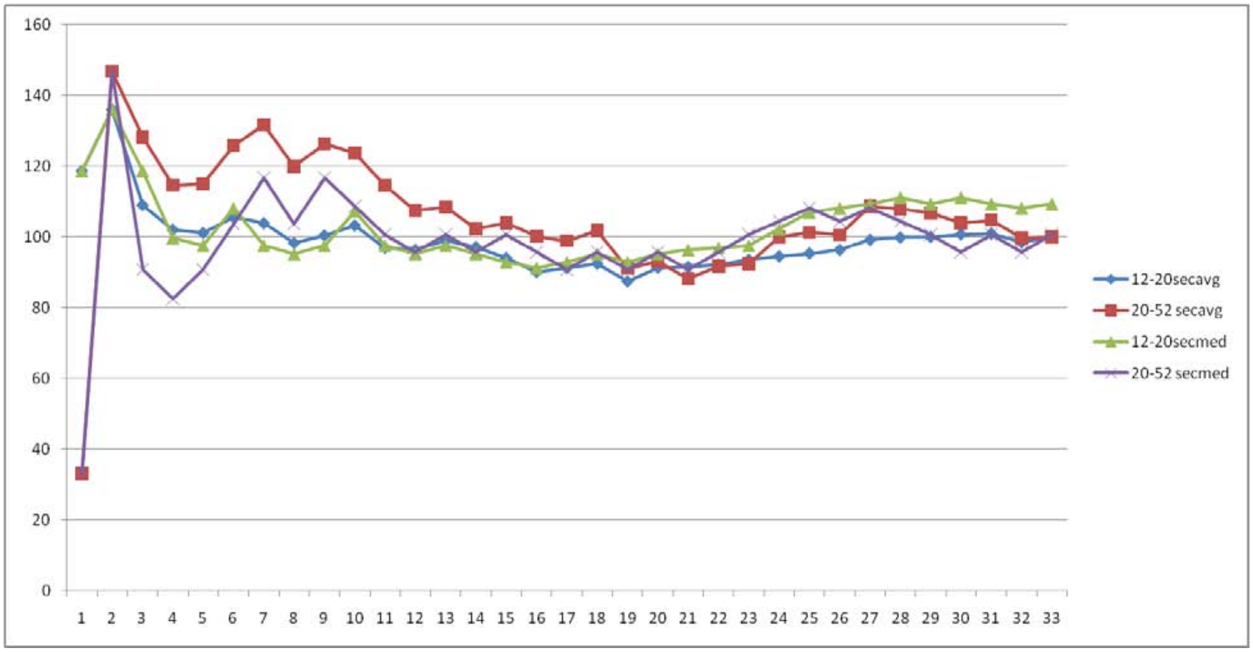


Figure 1.4

BA-4 Convergence Properties for the Mean and Median for the 12-20 Second And 20-52 Second Sampling Intervals

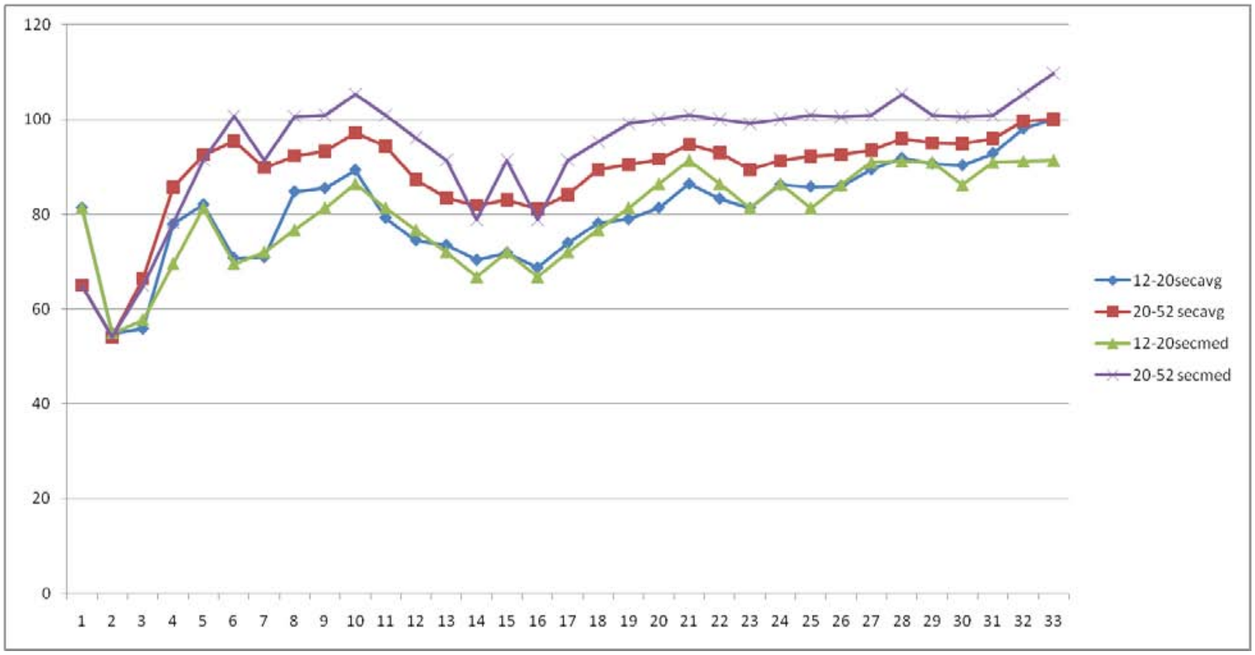
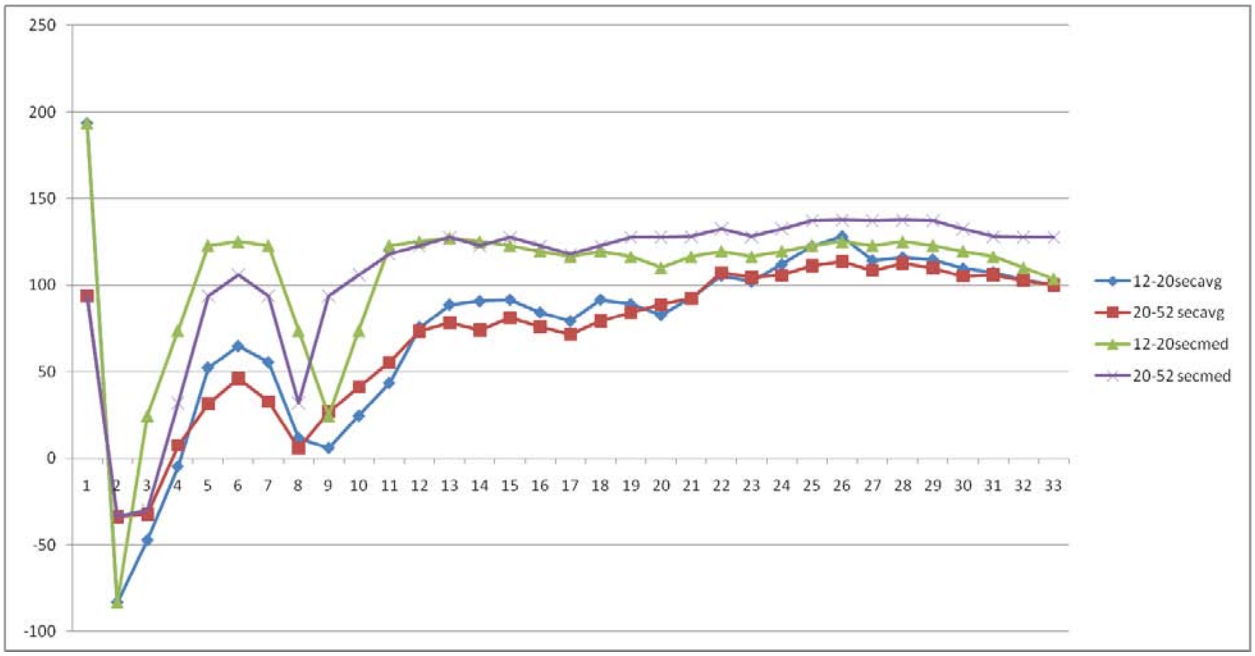


Figure 1.5

BA-5 Convergence Properties for the Mean and Median for the 12-20 Second And 20-52 Second Sampling Intervals



Tables 9.1 through 12.2 have been developed to explore the resiliency of the mean, median, and linear regression averaging techniques to the inevitable data contamination problems found commonly in actual net interchange values. For BA1 and BA2, extensive analyses were performed to evaluate the nature of their data contamination. Time did not allow for the other BAs to be analyzed, but the results would seem to be representative given the negative value head count information shown for those BAs in Table 8.

In the development of Tables 9.1 and 9.2, subject matter experts established 4 criteria for evaluating the calculated value of each sample. Assuming a positive value for a sample shows frequency response that supports interconnection frequency, these criteria are:

- A value below which it is very unlikely for the sample to be a valid measurement, and its inclusion would result in an underestimate of the calculated frequency response
- A value below which the sample is clearly not credible, and its inclusion would result in an underestimate of the calculated frequency response
- A value above which it is very unlikely for the sample to be a valid measurement, and its inclusion would result in an overestimate of the calculated frequency response
- A value above which the sample is clearly not credible, and its inclusion would result in an overestimate of the calculated frequency response

Two key points in developing the low limit values noted above are (1) the knowledge that load will not increase in response to a decreasing frequency, and (2) a knowledge of generator response and its extremely low likelihood that generators in aggregate will reduce output more than they will increase output. On the high side, large values significantly different in magnitude from the other values is the key.

The 4 rows shown in Table 9.1 and 9.2 correspond to the head count of samples that meet those exclusionary criteria. The data is presented for 3 sampling intervals, and in four groupings: all 70 events, the first and second set of 25 events based on chronology, and the last 25 events which include the last 5 samples of the second set of 25 events.

In reviewing Table 9.1, BA1's samples have a strong plurality of contaminated samples that tend to underestimate frequency response in all sampling intervals when all samples are considered. And note the difference in contamination in the first versus second 25 event sample sets. BA1's data does not at all support a hypothesis that errors will balance out with 70 samples, and results could be drastically different within the targeted 25 sample groupings.

In reviewing Table 9.2, BA2's samples have a strong plurality of contaminated values that tend to underestimate frequency response for the 12 to 20 second metric, but the contamination level is balanced for the other sampling intervals. However, note the substantial directional difference between the first and second sets of 25 samples. Again the balancing hypothesis is not supported with the targeted 25 sample groupings. ***In summary, analysis of the symmetry of contaminated data performed thus far does not at all support an assumption that contamination will balance out for a 25 event sample size.***

Table 9.1

Head Count Of Contaminated Data Samples For Different Sample Groupings And Sampling Intervals For BA-1

# of contaminated samples BA1														
70 total samples available														
		measure 12-20 sec					measure 20-52 sec					measure 20-40 sec		
	all	1 to 25	26-50	46-70		all	1 to 25	26-50	46-70		all	1 to 25	26-50	46-70
low-very hi and hi confidence	18	4	12	6		16	3	9	8		19	4	11	8
low-very high confidence only	12	2	10	4		10	0	7	6		13	3	8	6
hi- very high confidence only	1	0	1	0		5	1	1	3		3	1	1	1
hi- very hi and hi confidence	2	1	1	0		7	3	1	3		7	3	1	3

Table 9.2

Head Count Of Contaminated Data Samples For Different Sample Groupings And Sampling Intervals For BA-2

# of contaminated samples BA2														
70 total samples available														
		measure 12-20 sec					measure 20-52 sec					measure 20-40 sec		
	all	1 to 25	26-50	46-70		all	1 to 25	26-50	46-70		all	1 to 25	26-50	46-70
low-very hi and hi confidence	31	15	8	9		14	6	4	4		16	7	4	5
low-very high confidence only	22	9	6	7		12	5	4	3		11	4	4	3
hi- very high confidence only	4	0	4	0		13	3	9	4		13	3	9	4
hi- very hi and hi confidence	8	1	7	2		14	3	10	5		15	4	9	5

Using the same sampling intervals and event groupings as Tables 9.1 and 9.2, Tables 10.1 and 10.2 were developed to show the change in the values for mean, median, and linear regression from the all events within the four groupings after the two levels of exclusion are applied. The frequency response value used for normalization within Tables 10.1 and 10.2 are shown in gold, which is the median of the full 70 sample set after all exclusionary criteria are applied.

Tables 11.1 and 11.2 were developed from Tables 10.1 and 10.2 by taking the absolute value of the change between calculated value of all events within the grouping value and the calculated value after the each level of exclusionary criteria are applied. The intent here is to determine which averaging technique will produce a value closer to its value in the absence of contamination, given that the contamination ***will be inevitably present*** when used by the industry on an annual basis. At the bottom of Tables 11.1 and 11.2, the sum of the absolute differences are computed and used as a ranking mechanism to choose among the averaging techniques.

In reviewing Table 11.1 for BA1, the median overwhelmingly outperforms the mean, being more resilient in all groupings except for the 46-70 sample set with the 12 to 20 second sampling interval, in which the mean is only slightly better. (It is an insignificant 1% better when in event grouping 1-25 with the 20 to 40 second metric with the highest level of exclusion only.) The overall ranking score reflects the superiority of the median versus the mean, even if group 26-50 is dropped as it is influenced by BA1 contingencies. The median also outperforms the linear regression method for resiliency to data contamination in all groupings.

Table 10.1

Comparison Of Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval Event Grouping, And Exclusionary Criteria For BA1

Normalized Frequency Response - BA1			measure 12-20 sec				measure 20-52 sec					measure 20-40 sec		
case	group	#samples	mean	median		#samples	mean	median	regress		#samples	mean	median	
all	1 to 70	70	21	66		70	27	75	51		70	25	74	
very hi confidence exclusion	1 to 70	57	89	87		55	82	80	77		54	91	87	
hi & very hi confidence exclusion	1 to 70	50	95	100		47	85	84	80		44	89	90	
all	1 to 25	25	86	96		25	112	92	102		25	107	92	
very hi confidence exclusion	1 to 25	23	95	99		24	88	92	86		21	100	100	
hi & very hi confidence exclusion	1 to 25	20	97	100		19	89	92	86		18	93	96	
all	26-50	25	-103	46		25	-116	54	-25		25	-114	49	
very hi confidence exclusion	26-50	14	72	59		17	68	60	63		16	68	70	
hi & very hi confidence exclusion	26-50	12	79	62		15	75	67	71		13	77	83	
all	46-70	25	64	72		25	70	74	51		25	66	74	
very hi confidence exclusion	46-70	21	92	103		16	79	77	72		18	98	91	
hi & very hi confidence exclusion	46-70	19	100	104		14	87	82	77		14	90	91	

Table 10.2

Comparison Of Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval Event Grouping, And Exclusionary Criteria For BA2

Normalized Frequency Response - BA2			measure 12-20 sec				measure 20-52 sec					measure 20-40 sec		
case	group	#samples	mean	median		#samples	mean	median	regress		#samples	mean	median	
all	1 to 70	70	36	62		70	109	112	85		70	97	110	
very hi confidence exclusion	1 to 70	44	98	96		45	110	111	112		46	110	110	
hi & very hi confidence exclusion	1 to 70	31	104	100		42	112	112	111		42	113	110	
all	1 to 25	25	46	52		25	110	89	107		25	109	96	
very hi confidence exclusion	1 to 25	16	89	91		17	106	101	110		18	105	99	
hi & very hi confidence exclusion	1 to 25	9	108	101		16	109	106	111		14	111	107	
all	26-50	25	40	97		25	146	132	123		25	125	145	
very hi confidence exclusion	26-50	13	109	98		12	113	109	110		12	118	112	
hi & very hi confidence exclusion	26-50	10	102	97		11	108	99	104		12	118	112	
all	46-70	25	42	82		25	87	117	41		25	74	126	
very hi confidence exclusion	46-70	18	109	110		18	115	115	116		18	111	118	
hi & very hi confidence exclusion	46-70	14	108	110		16	115	115	113		15	116	126	

Table 11.1

Change In Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval Event Grouping, And Exclusionary Criteria For BA1

BA-1 % change from full group size			measure 12-20 sec				measure 20-52 sec					measure 20-40 sec		
			delta	delta			delta	delta	delta			delta	delta	
case	group	#samples	mean	median		#samples	mean	median	regress		#samples	mean	median	
all	1 to 70	70				70					70			
very hi confidence exclusion	1 to 70	57	68	21		55	54	5	26		54	66	13	
hi & very hi confidence exclusion	1 to 70	50	74	34		47	58	9	29		44	64	16	
all	1 to 25	25				25					25			
very hi confidence exclusion	1 to 25	23	9	3		24	24	0	16		21	7	8	
hi & very hi confidence exclusion	1 to 25	20	11	4		19	23	0	16		18	13	3	
all	26-50	25				25					25			
very hi confidence exclusion	26-50	14	175	13		17	185	5	88		16	183	21	
hi & very hi confidence exclusion	26-50	12	183	16		15	191	13	96		13	191	34	
all	46-70	25				25					25			
very hi confidence exclusion	46-70	21	28	32		16	10	3	21		18	32	17	
hi & very hi confidence exclusion	46-70	19	36	33		14	17	8	26		14	24	17	
		totals	584	155			562	43	318			579	129	

Table 11.2

Change In Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval Event Grouping, And Exclusionary Criteria For BA2

BA-2 % change from full group size			measure 12-20 sec				measure 20-52 sec					measure 20-40 sec		
			delta	delta			delta	delta	delta			delta	delta	
case	group	#samples	mean	median		#samples	mean	median	regress		#samples	mean	median	
all	1 to 70	70				70					70			
very hi confidence exclusion	1 to 70	44	62	34		45	2	1	28		46	14	1	
hi & very hi confidence exclusion	1 to 70	31	68	38		42	4	0	27		42	17	0	
all	1 to 25	25				25					25			
very hi confidence exclusion	1 to 25	16	43	40		17	4	12	3		18	4	4	
hi & very hi confidence exclusion	1 to 25	9	62	50		16	1	17	5		14	3	11	
all	26-50	25				25					25			
very hi confidence exclusion	26-50	13	69	1		12	33	23	13		12	7	32	
hi & very hi confidence exclusion	26-50	10	63	0		11	39	32	19		12	7	32	
all	46-70	25				25					25			
very hi confidence exclusion	46-70	18	66	28		18	28	2	75		18	38	8	
hi & very hi confidence exclusion	46-70	14	65	28		16	28	2	73		15	42	0	
		totals	498	217			137	88	242			132	88	

In reviewing Table 11.2, the median outperforms the mean overall for all sampling intervals. The mean outperforms the median for the 20 to 52 second sampling interval for event grouping 1-25, and in the 20 to 40 second metric for event grouping 26-50 and for one level of exclusion in grouping 1-25.

For BA2, the median outperforms the linear regression for the full sample set and by a large amount in sample set 46-70. The linear regression value outperforms the median by only 9 to 13% in event groupings 1-25 and 26-50.

In summary, the median is projected to be more resilient than either the mean or linear regression in the inevitable presence of contaminated actual net interchange data, and should be the averaging technique of choice unless a larger data set is analyzed in the future and a different metric is determined empirically to be more resilient to data contamination.

The differences between mean, median, and regression frequency response calculations, before and after the removal of contaminated data are of interest. Tables 12.1 and 12.2 have been developed to illustrate these differences.

In reviewing Table 12.1 for BA1, once the data contaminants have been removed, the mean and median differ by less than 10% in 21 out of 24 cases shown, and by 5% or less in 18 out of 24 cases. Also, the regression and median differ by 10% or less in 8 out of 8 cases, and by 5% or less in 6 out of 8 cases.

In reviewing Table 12.2 for BA2, once the data contaminants have been removed, the mean and median differ by less than 10% in 22 out of 24 cases shown, and by 5% or less in 15 out of 24 cases. Also, the regression and median differ by 10% or less in 8 out of 8 cases, and by 5% or less in 6 out of 8 cases.

In summary, the mean, median, and regression will yield very similar results once contaminated data is removed for all sampling intervals and event groupings.

Table 12.1

Differences In Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval And Event Grouping, Before And After Applying Exclusionary Criteria For BA1

			measure 12-20 sec		measure 20-52 sec			measure 20-40 se
Table 12.1			mean		mean	regress		mean
BA-1 Differences From Median			minus		minus	minus		minus
case	group	#samples	median	#samples	median	median	#samples	median
all	1 to 70	70	-46	70	-48	-24	70	-49
very hi confidence exclusion	1 to 70	57	2	55	1	-3	54	4
hi & very hi confidence exclusion	1 to 70	50	-5	47	1	-3	44	-1
all	1 to 25	25	-10	25	20	10	25	14
very hi confidence exclusion	1 to 25	23	-4	24	-4	-7	21	0
hi & very hi confidence exclusion	1 to 25	20	-3	19	-3	-7	18	-2
all	26-50	25	-149	25	-171	-79	25	-163
very hi confidence exclusion	26-50	14	13	17	9	3	16	-1
hi & very hi confidence exclusion	26-50	12	17	15	8	3	13	-5
all	46-70	25	-8	25	-4	-23	25	-8
very hi confidence exclusion	46-70	21	-11	16	2	-5	18	7
hi & very hi confidence exclusion	46-70	19	-4	14	5	-4	14	-1

Table 12.2

Differences In Frequency Response Resulting from Mean, Median, And Regression Averaging Techniques As A Function Of Sampling Interval And Event Grouping, Before And After Applying Exclusionary Criteria For BA2

			measure 12-20 sec		measure 20-52 sec			measure 20-40 se
Table 12.2			mean		mean	regress		mean
BA-2 Differences From Median			minus		minus	minus		minus
case	group	#samples	median	#samples	median	median	#samples	median
all	1 to 70	70	-26	70	-4	-28	70	-14
very hi confidence exclusion	1 to 70	44	2	45	-1	1	46	1
hi & very hi confidence exclusion	1 to 70	31	4	42	0	-1	42	3
all	1 to 25	25	-6	25	20	17	25	13
very hi confidence exclusion	1 to 25	16	-2	17	5	8	18	6
hi & very hi confidence exclusion	1 to 25	9	6	16	3	6	14	5
all	26-50	25	-57	25	15	-8	25	-19
very hi confidence exclusion	26-50	13	11	12	5	2	12	6
hi & very hi confidence exclusion	26-50	10	6	11	8	5	12	6
all	46-70	25	-40	25	-29	-76	25	-52
very hi confidence exclusion	46-70	18	-1	18	0	1	18	-6
hi & very hi confidence exclusion	46-70	14	-2	16	0	-2	15	-10

A comparison of Tables 12.1 and 12.2, before and after data contamination is removed, shows that means differ from medians in the presence of data contamination by 20% or more in 13 out of 24 cases, and by 10% or more in 19 out of 24 cases. Also, linear regression differs from medians in the presence of data contamination 20% or more in 5 out of 8 cases, and by 10% or more in 7 out of 8 cases. Given that mean, median, and regression values are usually very close once data contamination is removed (see Tables 11.1 and 11.2), the median value in the inevitable presence of data contamination yields a better approximation of the all 3 averaging techniques' values when the contamination is removed than either the mean or linear regression in the inevitable presence of data contamination, based on the above empirical studies. ***In summary, based on the empirical data presently available, the median is the best averaging technique to use, but this should be checked again as more data becomes available for more BAs.***

The impact of supplemental regulation on the resultant frequency response was evaluated for one BA. Unlike other dynamic transfers used to move generation and or load into or out of BAs, where there is a physical location from which frequency response may emanate, supplemental regulation when implemented as a pseudo-tie does not map into a physical location. The percent change in frequency response due to the inclusion of supplemental regulation in the actual net interchange value is shown in Table 13. ***In summary, the sample set size is too small to be conclusive, but further research is warranted at a moderate priority.***

Table 13

Percent Change In Frequency Response For A

BA When Its Supplemental Regulation Is Included In Actual Net Interchange

		12-20 second	20-52 second
averaging		sampling	sampling
technique		interval	interval
mean		-9	7
median		-14	4

Summary

Due to workload and concern for data confidentiality, relatively few BAs participated in this analysis, and as a result few of the analyses are conclusive. Below you will find a summary, in a somewhat sanitized form for readability, of the points made above.

1. ***It is highly recommended that BA's not be allowed to provide corrections for changes in scheduled net interchange.***
2. ***It is conclusive that all BAs need to use a common methodology to avoid "cherry picking".***
3. ***It is conclusive that some AGC action will appear as frequency response with the use of the 20 to 52 second metric.***
4. ***While more analysis is needed for many more BAs, the 20 to 52 second metric will flag squelched response.***
5. ***While more data is needed for more BAs, there does not seem to be a need to create special metrics based on the peak period, nor should sampling be restricted to on peak periods as was done in prior practices.***
6. ***Frequency response measures are sensitive to the method used to choose time zero – largest is easier to administer, but more research is needed in this area.***
7. ***It seems that events with point B above 59.95 Hz will result in lower scores, but a larger set of BAs are needed to make any firmer quantitative statements. It may make sense to have the majority of frequency events having a point B below 59.95 Hz.***
8. ***More analysis is needed to evaluate the impact of time skew related to the point A sampling interval.***
9. ***Many BAs need to provide scan rate data from 60 seconds before to 90 seconds after the frequency event to support an analysis to determine the best (or maybe just the least evil) sampling interval.***
10. ***It is impractical to rely on a single sample to be above some threshold value, and frequency response values will be meaningful ONLY when used with some type of averaging technique, and the choice of averaging***

technique will impact the accuracy of the score, given the data quality problems associated with using actual net interchange values.

11. ***25 samples should be a sufficient sample set size, provided that the right choices are made for the sampling interval and the averaging technique.***
12. ***Analysis of the symmetry of contaminated data performed thus far does not at all support an assumption that contamination will balance out for a 25 event sample size.***
13. ***The median is projected to be more resilient than either the mean or linear regression in the inevitable presence of contaminated actual net interchange data, and should be the averaging technique of choice unless a larger data set is analyzed in the future and a different metric is determined empirically to be more resilient to data contamination. See Appendix 2.***
14. ***The mean, median, and regression will yield very similar results once contaminated data is removed for all sampling intervals and event groupings.***
15. ***Based on the empirical data presently available, the median is the best averaging technique to use, but this should be checked again as more data becomes available for more BAs.***
16. ***Further research is needed to determine in supplemental regulation using a pseudo-tie affects the accuracy of frequency response calculations.***
17. ***When a larger data set becomes available, experimentation with averaging techniques such as a trimmed mean or the discounting of samples more than x standard deviations from the mean should be attempted.***

Appendix 1

List Of Frequency Events

							mHz dev	mHz dev
DAY	**DATE**	**TIME**	**HE**	point A	point C	point B	A to C	A to B
Mon	1/5/2009	10:26:56	11	60.014	59.978	59.977	36	37
Tue	1/20/2009	11:15:12	12	60.006	59.972	59.968	34	38
Tue	1/27/2009	1:39:24	2	60.013	59.971	59.967	42	46
Sat	2/7/2009	11:47:04	12	60.017	59.981	59.981	36	36
Sun	2/8/2009	10:53:41	11	60.000	59.959	59.962	41	38
Wed	2/18/2009	4:52:36	5	59.970	59.927	59.937	43	33
Thu	2/19/2009	5:59:40	6	59.992	59.942	59.943	50	49
Sat	2/21/2009	18:42:08	19	59.988	59.958	59.955	30	33
Sat	2/28/2009	1:15:52	2	60.001	59.960	59.959	41	42
Sun	3/1/2009	19:44:52	20	59.990	59.946	59.944	44	46
Tue	3/10/2009	16:16:40	17	60.006	59.971	59.969	35	37
Fri	3/13/2009	22:52:08	23	59.982	59.943	59.946	39	36
Sat	3/14/2009	3:06:20	4	60.003	59.968	59.962	35	41
Sun	3/15/2009	5:25:44	6	60.001	59.943	59.943	58	58
Mon	3/16/2009	9:58:20	10	60.004	59.968	59.971	36	33
Mon	3/16/2009	13:56:32	14	60.002	59.963	59.969	39	33
Mon	3/23/2009	12:21:40	13	60.021	59.982	59.968	39	53
Tue	3/24/2009	22:28:00	23	60.004	59.945	59.947	59	57
Thu	3/26/2009	4:51:36	5	60.004	59.917	59.929	87	75
Thu	3/26/2009	9:06:44	10	59.994	59.958	59.958	36	36
Sat	3/28/2009	21:21:36	22	59.997	59.958	59.954	39	43
Wed	4/1/2009	18:06:00	19	60.007	59.974	59.965	33	42
Sat	4/4/2009	14:45:08	15	59.998	59.965	59.962	33	36
Mon	4/13/2009	21:31:28	22	60.024	59.991	59.994	33	30
Wed	4/15/2009	14:11:24	15	59.989	59.937	59.942	52	47
Fri	4/24/2009	12:41:24	13	59.986	59.935	59.940	51	46
Sat	4/25/2009	23:03:04	24	59.973	59.916	59.923	57	50
Tue	4/28/2009	16:25:24	17	60.002	59.946	59.949	56	53
Sun	5/3/2009	11:06:00	12	60.012	59.956	59.955	56	57
Wed	5/6/2009	16:03:12	17	60.012	59.975	59.974	37	38
Wed	5/6/2009	22:57:08	23	60.019	59.964	59.942	55	77
Sat	5/9/2009	23:02:32	24	59.999	59.956	59.953	43	46
Wed	5/13/2009	15:16:08	16	59.993	59.960	59.960	33	33
Thu	5/21/2009	17:36:04	18	59.993	59.954	59.946	39	47
Tue	5/26/2009	22:52:12	23	60.001	59.947	59.946	54	55
Wed	5/27/2009	19:04:36	20	59.991	59.949	59.947	42	44
Tue	6/2/2009	0:34:08	1	60.006	59.971	59.971	35	35
Mon	6/15/2009	13:54:42	14	59.993	59.958	59.958	35	35
Thu	6/18/2009	11:01:04	12	59.990	59.958	59.959	32	31
Thu	6/18/2009	13:37:52	14	59.985	59.939	59.940	46	45
Sun	6/21/2009	17:50:52	18	59.999	59.946	59.947	53	52
Thu	6/25/2009	14:52:05	15	59.993	59.914	59.937	79	56

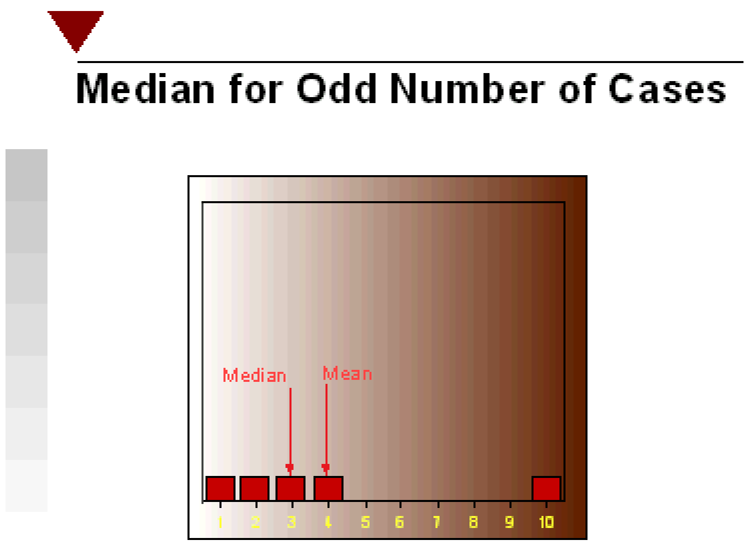
							mHz dev	mHz dev
DAY	**DATE**	**TIME**	**HE**	point A	point C	point B	A to C	A to B
Thu	7/2/2009	15:24:02	16	60.003	59.957	59.963	46	40
Thu	7/2/2009	22:44:21	23	60.015	59.939	59.984	76	31
Fri	7/3/2009	13:04:48	14	60.006	59.961	59.979	45	27
Mon	7/6/2009	14:35:49	15	60.006	59.954	59.949	52	57
Mon	7/13/2009	5:18:55	6	60.000	59.967	59.965	33	35
Sun	7/26/2009	9:20:30	10	60.009	59.968	59.967	41	42
Sun	7/26/2009	15:05:49	16	60.002	59.958	59.954	44	48
Mon	7/27/2009	15:36:00	16	59.996	59.966	59.968	30	28
Sat	8/15/2009	17:06:59	18	60.014	59.976	59.971	38	43
Sun	9/20/2009	11:38:21	12	60.013	59.981	59.979	32	34
Sat	9/26/2009	11:50:35	12	59.980	59.926	59.926	54	54
Sun	9/27/2009	4:14:38	5	60.022	59.992	59.991	30	31
Tue	9/29/2009	12:16:14	13	60.008	59.946	59.949	62	59
Fri	10/2/2009	6:52:26	7	60.004	59.965	59.966	39	38
Mon	10/19/2009	10:46:11	11	60.002	59.955	59.954	47	48
Mon	10/26/2009	15:53:57	16	59.993	59.936	59.932	57	61
Tue	10/27/2009	6:56:36	7	59.992	59.96	59.961	32	31
Mon	11/2/2009	11:09:41	12	59.976	59.946	59.941	30	35
Mon	11/2/2009	22:40:01	23	60.044	60.001	59.984	43	60
Tue	11/3/2009	20:42:08	21	59.997	59.966	59.946	31	51
Sun	11/15/2009	19:37:12	20	60.003	59.971	59.966	32	37
Wed	11/18/2009	22:37:50	23	60.018	59.987	59.985	31	33
Fri	11/20/2009	10:29:54	11	59.983	59.948	59.946	35	37
Tue	11/24/2009	12:53:50	13	59.98	59.949	59.948	31	32
Tue	11/24/2009	20:04:53	21	59.969	59.938	59.939	31	30
Sat	12/5/2009	10:03:30	11	59.976	59.944	59.947	32	29
Mon	12/7/2009	18:01:58	19	60.019	59.98	59.981	39	38
Tue	12/8/2009	8:43:58	9	59.998	59.962	59.964	36	34
Tue	12/8/2009	9:42:13	10	59.993	59.948	59.954	45	39

Appendix 2

Discussion Of Mean Versus Median

Median vs. Mean

Fechner also described relationship between the mean and the median in asymmetric distributions. The median, signified by the capital letter C, is the midpoint of an ordered series. When the scores are not equally distributed along the whole range of a variable, the median is likely a more appropriate measure of the central tendency than the mean. For instance, consider an ordered distribution of scores [1 2 3 4 10].



+++++

The Mean vs. the Median

As measures of central tendency, the mean and the median each have advantages and disadvantages. Some pros and cons of each measure are summarized below.

- ✍ The median may be a better indicator of the most typical value if a set of scores has an **outlier**. An outlier is an extreme value that differs greatly from other values.
- ✍ However, when the sample size is large and does not include outliers, the mean score usually provides a better measure of central tendency.

++++

In any skewed distribution (i.e., positive or negative) the median will always fall in-between the mean and the mode. As previously discussed in the section on "choosing an appropriate measure of central tendency", when dealing with skewed distributions, researchers typically decide between the mean or median as the best estimate of central tendency. As distributions go from symmetrical to more skewed, the researcher is more likely to chose the median over the mean.

++

When not to use the mean

The mean has one main disadvantage: it is particularly susceptible to the influence of outliers. These are values that are unusual compared to the rest of the data set by being especially small or large in numerical value. For example, consider the wages of staff at a factory below:

Staff	1	2	3	4	5	6	7	8	9	10
Salary	15k	18k	16k	14k	15k	15k	12k	17k	90k	95k

The mean salary for these ten staff is $30.7k. However, inspecting the raw data suggests that this mean value might not be the best way to accurately reflect the typical salary of a worker, as most workers have salaries in the $12k to 18k range. The mean is being skewed by the two large salaries. Therefore, in this situation we would like to have a better measure of central tendency. As we will find out later, taking the median would be a better measure of central tendency in this situation.

Another time when we usually prefer the median over the mean (or mode) is when our data is skewed (i.e. the frequency distribution for our data is skewed). If we consider the normal distribution - as this is the most frequently assessed in statistics - when the data is perfectly normal then the mean, median and mode are identical. Moreover, they all represent the most typical value in the data set. However, as the data becomes skewed the mean loses its ability to provide the best central location for the data as the skewed data is dragging it away from the typical value. However, the median best retains this position and is not as strongly influenced by the skewed values. This is explained in more detail in the skewed distribution section later in this guide.

have to take the 5th and 6th score in our data set and average them to get a median of 55.5.

++++

Median

- ? A commonly used robust and resistant measure of central tendency.
- ? Defined as the middle value when observations are ordered from smallest to largest.
- ? Divides the dataset into two parts of equal size, with 50% of the values below the median and 50% of the values above the median.
- ? Also known as the 50th percentile.
- ? Insensitive to extreme values.